

DRAFT

IMPROVING ACCURACY AND REDUCING COSTS OF
ENVIRONMENTAL BENEFIT ASSESSMENTS

Volume VI

Improving the Accuracy of Hedonic Price Methods:
Econometric Analysis of Existing Data Sets

CENTER FOR ECONOMIC ANALYSIS
UNIVERSITY OF COLORADO
BOULDER, CO 80309

September, 1985

DRAFT
September 1985

IMPROVING ACCURACY AND REDUCING COSTS OF
ENVIRONMENTAL BENEFIT ASSESSMENTS

Volume VI

Improving the Accuracy of Hedonic Price Methods:
Econometric Analysis of Existing Data Sets

by

Phil Graves
Jim Murdoch
Mark Thayer
Don Waldman

USEPA Contract #CR812054-01-1

Project Officers

Dr. Alan Carlin
Dr. Ann Fisher
Office of Policy Analysis and Planning
U.S. Environmental Protection Agency
Washington, D.C. 20460

OFFICE OF POLICY ANALYSIS AND PLANNING
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

DISCLAIMER

Although prepared with EPA funding, this report has neither been reviewed nor approved by the U.S. Environmental Protection Agency for publication as an EPA report. The contents do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

TABLE OF CONTENTS

	Page
I. Introduction	
A. Specification of the dependent variable	2
B. Degree of equilibrium versus disequilibrium in the housing market	3
C. Multi-market issues	6
D. Pollution variables	8
II. Data and preliminary regression results	11
A. The data	11
1. Air quality measures	11
2. Site variables	12
3. Neighborhood and community variables	12
B. Preliminary hedonic price gradient results	18
III. Variable selection	22
A. Sensitivity of VIS and TSP effects to included and excluded variables: the non-interacted case typical of the literature	22
B. Lot size interactions: disequilibrium issues	26
IV. Measurement Error	33
A. Introduction	33
B. Bounding the parameter estimates	33
1. Theory	33
2. An example	34
3. The methodology applied to our data	35
C. Minimum correlations	38
1. Theory	38
2. The methodology applied to our data	40
D. Conclusions	41
V. Effect of functional form	45
A. Introduction	45
B. The quadratic Box-Cox model	46
1. Specification	46
2. Estimation	47
3. Hypothesis testing	47
C. Empirical results	48
1. Specification 1	48
2. Specification 2	50
D. Conclusions	52
VI. Robust estimation	56
A. Introduction	56
B. The minimum absolute deviation estimator	57
VII. Concluding remarks	57
Bibliography	59

I. Introduction

Problems with the hedonic approach to environmental benefit estimation are becoming increasingly apparent (see, e.g., Atkinson and Crocker 1984, Bartik and Smith 1984, Berger, et al., 1984, and Horowitz 1984, Roback, 1982 and Rosen 1979). The problems are both theoretical and econometric. Each of the difficulties discussed here leads to a common issue: How accurate are estimated environmental benefits. While various studies are sometimes criticized as containing errors of commission or omission, until now there has been no systematic comparative analysis of the relative magnitude of potential errors.

In this component of the examination of the hedonic method we consider the robustness of marginal environmental values to (1) variable selection and treatment, (2) measurement error, (3) functional form, and (4) error distribution. One goal is to determine how sensitive property value-based estimates of marginal environmental values are to alternative assumptions. Another is to gain insights to guide future hedonic environmental benefit studies.

Section II describes the data set. No prior data sets have been able to explore the range of issues considered here.

Section III deals with variable selection, examining what difference alternative choices (and treatments) makes to the size and significance of the pollution effect on property values. A large body of literature is surveyed, from Ridker and Henning's 1967 study and continuing to the present. At the outset we should note that a very large number of analyses have been conducted which relate property values to air pollution, among other structural and neighborhood variables. The findings of these studies

show wide variation, in part because of variable definition and selection. Indeed, it is very difficult to find much commonality among these studies as will become clear.

In both of the two parts of Section III the variables are broken into three groups : “focus” variables (the pollution variables of policy interest), “sure” variables (variables known to exhibit a pronounced impact on property values), and “doubtful” variables (those whose uncertain inclusion or exclusion may or may not lead to instability of coefficients on the focus variables). Issues addressed at either the theoretical or empirical level are: 1) specification of the dependent variable, 2) degree of equilibrium versus disequilibrium in the housing market, 3) multi-market issues, and 4) the nature of the pollution variables. We turn briefly to each of these, prior to the more in-depth analysis of Section III.

A. Specification of the dependent variable.

A number of issues revolve around the specification of the dependent variable. The issue here is not transformation of the dependent variable although most authors take the logarithm of the dependent variable (see, e.g., Anderson and Crocker 1971, Brookshire, et. al. 1982, Murdoch and Thayer 1984, Peckham 1970, or Zerbe 1969). That issue is taken up in the later treatment of functional form; rather, we are concerned here with several other issues that have clouded the ability to compare existing contributions to the literature.

An organizing observation is that what would ideally be desired would be the site value of undeveloped land. This would eliminate the difficulties of controlling for structural traits and would reduce the need to control neighborhood traits if large open spaces were available for observation. Moreover, the bids for such parcels would presumably reflect

currently optimal development density--issues of disequilibrium or structural inertia would be of less importance. Unfortunately, the paucity of data on undeveloped parcels precludes such analyses. But what is the appropriate substitute?

Clearly, the second most preferred dependent variable would be actual sales transaction prices for individual homes. More recent studies have generally taken this approach (see, for example, Anderson 1981, Brookshire et. al. 1982, Goodman 1978, Brown and Pollakowski 1977, or Murdoch and Thayer 1984). However, the fact that most early studies, and some quite recent studies, were limited to census tracts or blocks (e.g., Ridker and Henning 1967, Zerbe 1969, Peckham 1970, Anderson and Crocker 1971, Nelson 1978, or Blomquist and Worley 1981) makes comparison of the impact of environmental quality difficult. The conclusions from such studies are suspect in that they stem from aggregation and employ owner-appraised values. Moreover, data on structural traits are typically weak, although a large and interesting range of neighborhood traits have been included. Although Kish and Lansing (1954) have argued that owner-assessment does not differ markedly from the assessments of professional appraisers, it is far from clear that pollution impacts would be unbiased. Other studies have employed multiple listing data or surveys (see Halvorson and Pollakowski (1981) for an example of the latter) which further reduce the ability to compare results among studies.

B. Degree of equilibrium versus disequilibrium in the housing market.

As already indicated, the hedonic literature implicitly assumes that the housing price variation observed represents equilibrating processes. That is, it is presumed that competition among identically-situated individuals

(similar tastes, income, and so on) causes any spatial advantages such as environmental quality to be bid into rents or property values so that there is no consumer surplus associated with a particular location. One fairly obvious, and frequently noted, implication of this is that if the number of people heavily damaged by pollution is small relative to the number of relatively clean locations, their high valuations of environmental quality will not be picked up in property value differentials. They will retain as consumer surplus their net benefits of occupying the cleaner sites.

The more important problem, and one that is pursued in the empirical specifications of Part B of Section III, is that the degree of equilibrium versus disequilibrium in the housing market has bearing on how “lotsize” is to be properly treated in the modeling. If the housing market is in short-run equilibrium only, then it is appropriate to enter lotsize as just another trait among the independent variables. This is, in fact, what is done in virtually all of the hedonic literature. The sole exception, Wieand 1973, does not make terribly clear why he proceeded as he did. To clarify the nature of the problem, consider the standard urban rent gradient model in which access to the employers located in the Central Business District is the only amenity in an otherwise featureless plain. In full (long-run) equilibrium, the gain in travel cost savings as one moves inward from the city’s edge is exactly offset by the higher rents per square foot of the high-access location. But since one receives the benefits of access regardless of how large a lot is purchased, the following equation characterizes the spatial equilibrium:

$$(1) dR \cdot L = dD \cdot C$$

Where dR = the rent differential

L = lotsize

dD = the change in distance (in miles)

C = cost of travel (per mile)

This equation captures the idea that the higher rents on the smaller lots which would be optimal at those higher rents must exactly offset the travel cost advantages giving rise to the rents. This equation can be rewritten to more closely approximate the standard urban model as in (2) below:

$$(2) \quad dR/dD = C/L$$

By then specifying how lotsize varies with rent (usually taken to be a unitary constant price elasticity), this model gives rise to the negative exponential rent gradient commonly employed by urban economists.

In the environmental application, as with travel costs, the benefits of clean air are dependent on where one is located but not on how large a lot one purchases. The value of clean air is a certain amount whether one is on a small lot or a large lot--if a large lot is purchased, then one is paying more for environmental quality than if a small lot is purchased. This is not true for structural traits (e.g. a swimming pool of constant size costs no more on a relatively small lot than it does on a large lot), but is true for all amenities having a public good nature, such as environmental quality. Hence, if the urban economy is characterized as being in long-run equilibrium, then all public good amenities (environmental quality but also all included neighborhood traits) should be interacted with lotsize. Illustrating, if a particular level of environmental quality is worth \$100 to an individual who happens to occupy an average-sized lot (say, 1/8th acre), then if another individual chooses

to purchase a full acre lot he or she must be paying \$800 for that level of environmental quality. Similar comments apply to school quality, crime, and other traits which are location-specific but not dependent on lotsize.

Whether the urban economy is in long-run or only short-run equilibrium is an empirical matter. For the case of pollution, it is likely that for certain cities housing density (the inverse of lotsize) has had time to adjust to long-term pollution patterns. Los Angeles, for specificity, has an airshed which is systematically dirtier in the more eastern locations and this has been the case for a long time. One would expect that the lower rents in these dirtier locations will have resulted in larger than average lotsizes. In Part B we allow for interactions between public good amenities and lotsize, carrying both the standard and the modified equations through the remaining robustness exercises.

C. Multi-market issues

It is increasingly being realized that spatial variations in amenities are not capitalized exclusively in either the land or labor markets (see Graves and Knapp 1985 for an intuitive treatment, Roback 1982, Haurin 1982, or Rosen 1979 for models of varying level of generality). The reason they have been so treated stems from the separation of the economic subdisciplines which have been involved in amenity valuation. Urban economists have viewed land capitalization of amenities as a natural extension of the basic theoretical model of urban economics, namely, the Alonso 1964-Muth 1969 rent gradient model discussed above. Travel cost advantages were presumed capitalized into rents, hence why should variation in other amenities not also be so capitalized? Labor economists meanwhile were arguing that average differences in amenity levels among urban areas

would be capitalized into wages; the idea was simply that areas which were desirable for whatever reason would reach an equilibrium characterized by lower wages than elsewhere. The implicit presumption which allowed such studies to proceed in isolation, was that labor markets were national while housing markets were local in nature. To be charitable, it might at first blush seem plausible that one might first select a city, considering only wages, and then select a location within that city based upon the structure of rents.

Yet, it is clear that rational households will not operate in that manner; rather, they will look jointly at wage rates, cost-of-living (which is principally related to local rents) and any amenities which are not fully capitalized in either market (representing disequilibrium influences). That someone might, for example, turn down a job in Detroit offering a somewhat higher wage than he or she was currently receiving in Santa Fe would surprise very few people. All one need suppose to generate interactions in capitalization between the land and labor markets is that people consider housing costs in their decision to move between labor markets; it would be implausible to assume otherwise. Hence, the urban model underestimates the value of access because the small town with high access (and low rents) also offers lower wages--the access value is captured in both markets. Similarly, the labor model underestimates the amenity value of, say, San Francisco's scenic bay because not only are wages lower, but also rents are higher than would be the case in the absence of the bay.

The preceding argument has important implications for valuations of environmental quality. First, the many property value studies systematically undervalue environmental benefits by ignoring simultaneous

wage capitalization as indicated above.

Second, there is no reason to suspect that the shares of the compensation for environmental quality which occur in the land and labor market should be constant across urban areas. That is, land scarcity might cause 90 percent of the compensation for air quality to occur in land markets in San Francisco, while in Phoenix perhaps only 40 percent of compensation would be in land markets with relatively more occurring in labor markets. This suggests that there is no reason why studies conducted in different cities should be comparable; indeed, results may be expected to be different, partly because of topographical differences, partly because of people differences (higher variance in income, for example, will lead to a different rent gradient for access in one city relative to another and similar results should hold for environmental quality!. The point here is that different cities have different supply functions and demand functions for all traits, including environmental quality, and, consequently, the hedonic function which is a reduced form can hardly be expected to be comparable among cities. This would imply that the proper functional form (discussed in Section IV) will vary among urban areas and the results presented here for Los Angeles are unlikely to be generalizable.

D. Pollution Variables

As with prior issues, the choice of focus variable is rather critical and the many studies vary greatly in how this choice is handled. To be properly reflected in property values an amenity must have two essential properties: 1) Individuals must know how it enters their utility functions, in much the same way that they can perceive the utility of a can

of Campbell's Cream of Mushroom soup prior to consuming it, and 2) they must further know how pollution varies in the feasible set of locations. Since even those who have studied the health, property damage, and damage, and aesthetic effects of pollution for large portions of their lifetimes have only a vague and very uncertain notion of the size of such effects, the first condition may appear dubious to many. We are not particularly troubled by this, however, as people routinely make many complicated judgements in deciding how important such things as views, humidity, and the like are to them. As economists, we have a bias toward assuming people perceive the impact that different levels of goods will have upon them.

The second condition, that people know the spatial array of pollution, is more problematical. Some pollutants, such as ozone or CO are odorless, colorless, and tasteless in ambient concentrations. Any property value effects found in the various studies for these, and similar pollutants, are likely to be due to correlations with other, observable pollutants. This suggests that perhaps TSP or visibility are more appropriate for inclusion in an hedonic equation than are the less observable pollutants.

Other problems relating to the focus variables are whether dispersion model values should be used rather than monitored values. Interpolation, however sophisticated, introduces measurement error. Another difficulty in comparing existing studies is the degree of "contemporariness" of the pollution data with other data in the analysis--this is particularly problematic for those early studies employing census data.

Section IV building on the preferred models of Section III, analyzes the issue of whether the environmental variables of interest can ever be

accurately portrayed, concluding that maximum likelihood estimators will generally be non-unique. Given priors on the minimum and maximum correlation between true and measured variables, parameter estimator ellipses are calculated giving a more accurate picture of the potential variability of these marginal effects. The procedure follows Klepper and Learner 1984, and has been used in a very similar manner by Atkinson and Cracker 1984.

In Section V, functional form is considered, with the opening observation being that economic theory usually has little to say about correct functional form. Since the hedonic equations are reduced forms stemming from several structural equations, it would appear to be unreasonable to impose a priori restrictions on the structure of the data. Yet, in nearly all studies to date (with the exception being Halverson and Pollakowski, 1981) there is, at best, a non-systematic search over a few functional forms, with the results reported being those most closely conforming to the priors of the investigator. We employ here a model sufficiently general (the quadratic Box-Cox) to include the most popular forms of the literature (linear, log-linear, log-log, quadratic and translog) as special cases. The nested nature of the various models enables conventional hypothesis tests to be conducted.

We turn in Section VI to a test of the robustness of environmental valuations, to alternative assumptions regarding the error distribution. Heavy reliance on assumptions of normality can lead to biased parameter estimates when, as many applied econometricians suspect, the true error distribution has greater weight in its tails. We employ in this section a more robust estimator than least square, the minimum absolute deviation estimator.

Section VII concludes this component of the broader research effort devoted to improving the accuracy and reducing costs of 'environmental benefit assessments.

II. Data and Preliminary Regression Results

A. The data

The data base, constructed to estimate a hedonic price equation, includes observations from Los Angeles, Orange, Riverside, and San Bernadino counties. The dependent variable in the analysis is the sale price of owner occupied single family residences. The amenity variable of interest is urban air quality. Four variables were used to represent air quality: (1) actual visibility; (2) total suspended particulate concentrations; (3) nitrogen dioxide concentrations; and (4) annual days exceeding the ambient ozone standard. These four separate measures were used to account for each component of perceived air quality (aesthetics, physical damage and health). The following discussion details further the air quality "focus" variables, the site variables, and the neighborhood and community variables.

1. Air quality measures

The visibility (capturing aesthetics) data consist of prevailing visibility recordings made by weather station personnel at airports and other weather stations (see Figure II.1 for a map of the stations). At each of these stations, median visibility was calculated from three observations per day over the two year period (2190 observations at each station). These were used to construct "isovisibility" contours for the study area (See Figure 11.2). A grid system was then developed to identify the median visibility at each location. The isovisibility

contours did not necessarily correspond to community or census tract boundaries; hence, the visibility data vary within communities and census tracts.

The remaining air quality data was taken from California Air Resources Board publications. Annual averages at each monitoring station for total suspended particulates (physical and health damage), nitrogen dioxide and days exceeding the ozone standard (health and material damages aspects) were converted to the grid system so that each location could be assigned relevant values. This procedure is superior to the more aggregated assignments of pollution to sites commonly employed in the literature.

2. Site variables

In addition to the air quality variables, the independent variable set includes variables which correspond to three levels of spatial aggregation: site, neighborhood and community. The site characteristic data were obtained from the Market Data Cooperative (a computerized clearinghouse for housing data) and pertain to homes sold in the third quarter of the 1979 time period. A large random sample of approximately 1400 observations was taken from the original data set of over 100,000 observations. The site characteristic data is at the household (micro) level and contains information on nearly every important structural and/or quality attribute. Included in the list of variables are those that pertain to both quantity (e.g., total number of rooms, square footage of living area, number of bathrooms) and quality (e.g., pool, view, number of fire places, air conditioning) of each particular house.

3. Neighborhood and community variables

Other variables which could significantly affect a home's sale price are those that reflect the condition of the neighborhood and

community in which it is located. In order to capture those impacts and to isolate the independent influence of location vis-a-vis air quality differences, several neighborhood and community variables were included in the econometric modeling. Neighborhood refers to surrounding census tract and includes variables such as income, ethnic composition, distance to work, and distance to the beach. Information from the 1980 census was utilized. Given the large number of census tracts (over 1,500 in the study area) variation in these data is quite substantial. Pertinent community (city level) variables include density measures, lot size, school quality, crime rate, and distance to the central business district. Also included are a set of zero-one dichotomous variables for county (Los Angeles, Orange, Riverside, San Bernadino) location. In contrast to the site and neighborhood characteristics there are only a limited number of communities, with correspondingly less variation. The data are completely described in Tables II.1 and II.2.

TABLE II.1

Variables Used in Analysis of Housing Market for 1978-79

n (hy hesized ousin ale pri			
<u>Dependent:</u>			
Sale Price	Sale price of owner occupied single family residences	(\$100)	Market Data Cooperative
<u>Independent-Site</u>			
Age	Age of home (negative)	Years	Market Data Cooperative
Bathrooms	Number of bathrooms (positive)	Number	Market Data Cooperative
Living Area	Square feet of living area (positive)	Square Feet	Market Data Cooperative
Pool	1 if pool, 0 if no pool (positive)	0=no pool, 1=pool	Market Data Cooperative
Fireplaces	Number of fireplaces (positive)	Number	Market Data Cooperative
View	1 if view present, 0 if not (positive)	0=no view, 1=view	Market Data Cooperative
Air Conditioning	1 if air conditioned, 0 if not (positive)	0=no a/c, 1=a/c	Market Data Cooperative
<u>Independent-Neighborhood:</u>			
Distance to Beach	Miles to nearest beach (negative)	Miles	Calculated
Ethnic Composition	Percent white in census tract (positive)	Percent	1980 Census
Time to Work	Average time to employment from census tract (negative)	Minutes	1980 Census
Income	Mean Income (positive)	\$	1980 Census
<u>Independent-Community:</u>			
School Quality	Community's 12th grade math score (positive)	Percent	California Assessment Program (1979)

TABLE II.1
(Continued)

Variable	Definition (hypothesized effect on housing sale price)	Unit	Source
<u>Independent-Community:</u> (Continued)			
Miles to Central Business District	Distance from census tract to dominant city in county (negative)	Miles	Thomas Brothers Grid Maps
Crime	Seven major crimes per 1000 people in surrounding communities (negative)	Crimes/persons	Summary Characteristics 1980 Census
Lot Size	Average size of residential lot in community (positive)	Square Yards	Calculated
County Location			
D1	Orange County Location (negative)	0 if no, 1 if yes	Thomas Brothers
D2	Riverside County Location (negative)	0 if no, 1 if yes	Grid Maps
D3	San Bernadino County Location (negative)	0 if no, 1 if yes	
<u>Independent-Air Quality:</u>			
Visibility	Airport readings of visual range (positive)	Median miles	Trijonis, et al (1984)
Total suspended Particulates	Particulate concentrations (negative)	Mg/m ³	California Air Resource Board
Nitrogen dioxide	Concentrations of nitrogen Dioxide (negative)	PPHM	California Air Resources Board
Days Exceeding Ozone Standard	Days per year that location exceeds federal ambient standard for ozone (negative)	Days	California Air Resource Board

TABLE II.2

Summary Statistics for Variables Used in the
Accuracy Analysis of the Hedonic Methodology

Variable	Mean	Standard Deviation	Minimum	Maximum
<u>Dependent:</u>				
Price (\$100's)	1033.14	574.38	200.00	4980.00
<u>Independent - Site:</u>				
Age	22.36	18.11	0.00	80.00
Bathrooms	1.90	.66	1.00	6.00
Living Area	1542.15	588.91	63.00	5890.00
Pool	.16	.37	0.00	1.00
Fireplaces	.76	.62	0.00	4.00
View	.07	.25	0.00	1.00
Air Conditioning	.56	.50	0.00	1.00
<u>Independent - Neighborhood:</u>				
Distance to Beach	12.62	8.22	.25	42.50
Ethnic Composition	85.57	12.42	5.40	99.10
Time to Work	23.52	3.22	15.00	36.00
Income	27889.82	12105.97	8738.00	115522.00
<u>Independent - Community:</u>				
School Quality	63.29	15.99	54.10	81.00
Miles to Business District	7.36	7.72	0.00	31.00
Crime	.05	.076	.02	1.67

TABLE II.2
(Continued)

Summary Statistics for Variables Used in the
Accuracy Analysis of the Hedonic Methodology

Variable	Mean	Standard Deviation	Minimum	Maximum
Lot Size	1978.22	7986.21	480.25	183350.84
County Location				
Orange	.34	.48	0.00	1.00
Riverside	.006	.08	0.00	1.00
San Bernadino	.04	.19	0.00	1.00
<u>Independent ~ Air Quality:</u>				
Visibility	11.06	2.32	7.00	15.00
Total Suspended Particulates	91.86	12.34	64.00	130.00
Nitrogen Dioxide	10.22	2.49	6.00	14.00
Days Exceeding Standard	92.46	53.67	12.00	197.00

B. Preliminary hedonic price gradient results

The estimated hedonic price gradients (linear and semi-log) which serve as the preliminary results are presented in Table II.3. A number of interesting aspects of the equations are readily apparent. First, the independent variable set was chosen to reflect the many characteristics of a home. Thus, living area and number of bathrooms represent the quantity of the home, whereas house age, or the presence of a pool, fireplaces, view and air conditioning relate to quality. In addition, characteristics which reflect the immediate neighborhood (ethnic composition), location (time to work, distance to beach), and surrounding community (lot size, distance to business district) are included. Finally, two measures of air quality visibility and total suspended particulates complete the independent variable set. Additional measures of air quality were not included to prevent collinearity in addition to the reasons discussed earlier. It should be noted at this point that the specification presented is only one of many possible models. The impact of various included/excluded variables is the concern of the next section.

The second noteworthy aspect of the equations is that the non-linear specification outperforms the linear specification. This is consistent with the conjecture of Rosen (1974) who noted that consumers cannot always arbitrage by dividing and repackaging housing attributes. Third, a significant portion (.69-.73) of the variation in home sale price is explained by the independent variable set. Fourth, in each equation ten of the fourteen (excluding the constant) estimated coefficients are significantly different from zero at the five percent level. The exceptions are air conditioning, pool distance to work, lot size (linear)

unexpected manner, that is, their relationship to the dependent variable is contrary to prior expectations.

The final aspect worth noting is that the air quality variables are significantly different from zero and possess the expected relationship to home sale price. These results imply that individuals are acting upon air quality information when making locational choices. The monetary impact of a one unit change in visibility (total suspended particulates) ranges from \$6818 (\$1239) to \$8767 (\$1542), dependent upon functional form.

Given those preliminary results the importance of including or excluding other variables, measurement error, functional form and the distribution of the error terms is analyzed in the following sections.

TABLE II.3

Preliminary Hedonic Price Gradient Estimates
(t - ratios in parenthesis)

Dependent Variable = Home Sale Price (\$100)

Variable	Linear	Semi-log
Age	1.938 (3.55)	.0003 (.70)
Bathrooms	61.88 (21.21)	.043 (2.93)
Living Area	.536 (22.03)	.00038 (22.24)
Pool	28.47 (1.20)	.011 (.67)
Fireplaces	47.45 (2.83)	.083 (7.08)
View	422.37 (12.16)	.214 (8.86)
Air Conditioning	-9.90 (-.48)	-.014 (-.98)
Distance to Beach	8.562 (5.07)	.006 (5.04)
Ethnic Composition	4.59 (6.25)	.006 (11.81)
Time to Work	-2.04 (-.69)	-.0003 (-.16)
Distance to Business District	5.82 (4.00)	.003 (3.17)

TABLE II.3
(Continued)

Preliminary Hedonic Price Gradient Estimates
(t- ratios in parenthesis)

Dependent Variable = Home Sale Price (\$100)

Variable	Linear	Semi-log
Lot Size	.001 (1.21)	.000002 (2.14)
Visibility	87.67 (14.14)	.066 (15.35)
Total Suspended Particulates	-16.42 (-15.29)	-.012 (-15.63)
Constant	22.48 (.20)	5.83 (74.12)
R-Square	.698	.737
Sum of Square Residuals	140978117	68.296

III. Variable Selection

We now turn to an analysis of how sensitive the estimated property value effects of pollution are to variable selection. As indicated at the outset, the potential variables for the hedonic price equation are divided into “focus” variables, visibility (VIS) and total suspended particulates (TSP), “doubtful” variables (INCCT, CRME, SCHOOL, and D1 -D3), and “sure” variables (AREAS, BATH, AGE, LOTSZ, FIRE, POOL, WHITCT, WRKCT, BEACH, CBD, and the constant).

A. Sensitivity of VIS and TSP effects to included and excluded variables : the non-interacted case typical of the literature

In this subsection, the estimated coefficients on the focus variables, VIS and TSP, are examined when different combinations of them and of the doubtful variables are entered into the hedonic equation.* In addition, some attention is given to the coefficients on the other included variables. The results of this inquiry suggest that variable selection can dramatically affect the estimate of the relationship between home prices and environmental quality.

The doubtful variables were chosen primarily for illustration. However, INCCT, CRMS, and SCHOOL have been included and excluded in previous studies indicating some uncertainty about their relationship to home prices. Income constrains the consumer, hence determines the choice of greater levels of characteristics. While many of the studies which

*As our earlier discussion of the choice of pollution variable suggested would be possible, initial runs including NO₂ and a measure of violations of ozone standards among the focus variables suggested that these pollutants do not have an impact on property values--NO₂ always had a positive relationship with home price (likely reflecting access to freeways) while the ozone measure had a coefficient which was highly variable and rarely significant.

included income were conducted prior to Rosen's 1974 two-stage procedure, even recent studies such as that of Halvorson and Pollakowski 1981 have entered census tract income as a proxy for unobserved neighborhood traits. However, this variable will be highly correlated with the income of the household, making it a doubtful variable from our perspective. Some researchers have argued that lower crime rates are expected to increase house prices, while others have indicated that higher house prices attract criminal activity and mean greater crime rates. Thus CRME was placed into the doubtful variable category. Public school quality, apart from being difficult to measure, may not matter to the extent that those in higher priced homes have higher income and send their children to private schools. The Los Angeles study area covered three counties and we did not know how the net effect of the bundle of tax and publicly provided goods on property values would differ among them. Therefore, the county dummies were placed in the doubtful category.

The other included variables represent measures that have traditionally appeared in hedonic studies. They correspond to measures of site specific characteristics, neighborhood quality characteristics, and location parameters. Although not undertaken here, a more general analysis might consider all independent variables, or at least a larger subset than that selected here, as doubtful. However, the size of the independent variable set would make this approach computationally intimidating-- moreover, the general conclusions could only be strengthened.

The estimated coefficients on the focus variables as they vary with inclusion or exclusion of doubtful variables are presented in Table III-3. Consider first the results for VIS (See Columns 1 and 3 of Table III-1).

VIS appears to be particularly sensitive to the county dummies. Without TSP in the equations, the estimates of the VIS effect on property values range from positive and significant (in the sense that the t-ratio exceeds 2.00) to negative and significant when the county dummies are entered. When TSP and VIS are included in the equation, the estimates on the effect of VIS go from positive and significant to insignificant. This is an important finding because it illustrates how “fragile” an inference about the relationship between VIS and PRICE may be. This is possibly a finding unique to the study area and, without further information, we cannot determine if VIS measures a county influence or an environmental influence. However, Murdoch and Thayer (1984) found that the relationship remained positive and significant when a larger sample size was used.

For TSP, the signs of the coefficients are always negative and the t-ratios are always greater than 2.00. With only TSP in the equation, the coefficient estimates jump when the county dummies are entered. However, when VIS and TSP are entered together, the estimates for TSP exhibit remarkable stability, ranging from -.008 to -.012.

Of the other included variables, AREA, BATH, FIRE, POOL, and WHTCT are fairly stable and approximately within the range of estimates presented in earlier tables, and in the literature. The estimated effect of AGE goes from negative and significant to positive and significant, indicating that it is correlated with one or more of the doubtful variables. VIEW is always positive and significant with a range on the estimates of approximately .25 to .08. BEACH exhibits a negative and significant relationship when either VIS or TSP is entered into the equation.

TABLE III.1

Estimated Coefficients for TSP and VIS by Possible
Combinations of Doubtful Variables.

Dependent Variable = the Natural Logarithm of PRICE

Doubtful Variables ^a .	b		VIS and TSP together ^c	
	VIS only	TSP only		
none	.030	-.006	.066	-.012
X ₁	.022	-.004	.050	-.009
X ₂	.036	-.005	.073	-.012
X ₃	.030	-.006	.066	-.012
X ₄	-.046	-.010	.004*	-.011
X ₁ X ₂	.024	-.004	.054	-.009
X ₁ X ₃	.022	-.004	.051	-.009
X ₁ X ₄	-.040	-.008	-.005*	-.008
X ₂ X ₃	.036	-.005	.074	-.012
X ₂ X ₄	-.042	-.010	.004*	-.010
X ₃ X ₄	-.046	-.011	.004*	-.011
X ₁ X ₂ X ₃	.024	-.004	.054	-.009
X ₁ X ₂ X ₄	-.038	-.008	-.004*	-.008
X ₁ X ₃ X ₄	-.040	-.008	-.005*	-.008
X ₂ X ₃ X ₄	-.042	-.010	.004*	-.010
X ₁ X ₂ X ₃ X ₄	-.039	-.008	-.004*	-.008

a. The doubtful variables are: X₁ = INCCT, X₂ = CRME, X₃ = SCHOOL,
X₄ = DI, D2, D3.

b. The hedonic equation is estimated with only one focus variable

c. The hedonic equation is estimated with both focus variables.

* indicates that the t-ratio is less than 2.00.

However, when both pollution variables are entered, BEACH is sometimes positive and significant. The effect of BEACH is critically affected by the county dummies. This may indicate that BEACH, VIS, and the county dummies are measuring a similar influence. Earlier results indicated that these were also sensitive to AIR. This result did not hold true as the sample size was increased (see Murdoch and Thayer 1984). WRKCT and CBD both exhibit effects which go from negative to positive as various doubtful variable combinations are used.

B. Lotsize interactions: disequilibrium issues

As indicated in the introductory discussion, it will be appropriate to interact public good amenities with lotsize if the land market is in long run equilibrium. The argument was that the locations with, for example, clean air will tend to rent or sell for more, but the higher land prices will lead to economizing on land in the production of housing (since the value of exposure to clean air is independent of how large a lot is purchased). The extent to which the land market in Los Angeles is in equilibrium or disequilibrium is, of course, an empirical matter.

We explore in this section three possibilities: 1) the land market is only in short-run equilibrium (air quality entered but not interacted), 2) the land market is in full long-run equilibrium (air quality is entered only in interaction with lotsize and 3) the land market is not in full long-run equilibrium but partially reflects long-run equilibrium considerations (air quality is entered separately as well as being interacted with lotsize). The partial effect of air quality on property values will generally vary according to the treatment of lotsize and a natural question is: How robust are marginal environmental evaluations to lotsize treatment?

Tables III.2 and III.3 display the regressions corresponding to the two unexplored possibilities above, retaining the distinctions on specification and included/excluded doubtful variables. Table III.2 presents results for the long-run equilibrium view (public good amenities only entered in interacted form) while Table III.3 enters the public good amenities both separately and interacted with lotsize. There is, of course, considerable information to be gleaned from these tables in comparison to our earlier base case (Table II.3). We emphasize here only the impact on the air quality variables.

In Table III-4 the coefficients on the air quality variables are reproduced from equations presented earlier. Columns 1 and 2 are the two linear specifications, while the semi-log results are shown in columns 3 and 4. Within each of these specifications are three sub-columns representing the three possibilities discussed above, containing coefficients from Tables II.3, III.2, and III.3. At the foot of this table are reproduced the means of lotsize and price which are necessary to calculate the effect (at the means) of air quality on property values.

In Table III-5 these effects are calculated and the appropriate degree of disequilibrium is seen to matter greatly. VIS is seen to range across possibilities from plus \$17,347 to -\$7,792, although the negative figures are not significantly different from zero. It would, however, appear that the positive effect of greater visibility is highly variable across both specification and interactive treatment. Note that differences between effects for possibility one and possibility two are quite close in three of four cases.

The situation for TSP is even more disturbing for those interested in valuing pollution effects--significant effects range from large and positive (see the fully interacted results with either the linear 1 or semilog 1 specification).

Hence, looking only at effects of lotsize interaction at the means of the data we find that the issue of equilibrium versus disequilibrium is both important and underinvestigated. Coefficients are not robust to alternative treatments of lotsize either within or across specifications. Since the degree of equilibrium will depend on many factors which vary across cities, it is not surprising that results, regardless of specification, vary among data sets from many cities.

TABLE III.2

Preliminary Regression Results for the
Linear and Semilog Functional Forms

(t-ratios is in parenthesis)

Variable	Linear		Semilog	
	Specification 1	Specification 2	Specification 1	Specification 2
AREA	.576 (22.02)	.548 (21.44)	.00039 (21.08)	.00038 (20.57)
BATH	84.315 (3.68)	112.381 (5.00)	.067 (4.05)	.082 (5.07)
AGE	2.036 (3.56)	.715 (1.22)	.00009 (.21)	-.0008 (-1.93)
LOTSZ	-.089 (-1.61)	.557 (5.96)	-.00016 (-3.99)	.0002 (3.04)
FIRE	26.578 (1.75)	33.398 (1.89)	.078 (6.02)	.081 (6.36)
AIR	-21.965 (-1.05)	-18.548 (-.90)	-.029 (-1.92)	-.0211 (-1.42)
POOL	17.291 (.67)	31.850 (1.27)	.0076 (.41)	.0141 (.78)
VIEW*	.063 (9.51)	.059 (7.72)	.00005 (10.25)	.00004 (7.38)
WHTCT*	.0010 (4.39)	.00096 (4.01)	.0000011 (6.53)	.0000009 (5.40)
WRKCT*	-.00011 (-.10)	-.003 (-2.17)	-.0000006 (-.67)	-.0000012 (-1.18)
BEACH*	-.0013 (-2.40)	-.0088 (-5.70)	-.0000013 (-3.46)	-.000005 (-4.08)
CBD*	.006 (6.52)	.00068 (.61)	.0000046 (7.15)	.000001 (1.69)
VIS*	.044 (13.30)	.0097 (.61)	.000033 (13.95)	.00002 (3.69)
TSP*	-.0045 (-10.19)	-.0047 (-6.97)	-.0000027 (-8.31)	-.0000033 (-6.85)
CRME*		-.0056 (-.58)		.000007 (1.02)
D1*	-	-.201 (-8.43)	-	-.0001 (-6.83)
D2*		.116 (4.86)	-	.00005 (2.71)
D3*	-	.101 (3.74)	-	.00005 (2.36)
INTERCEPT	-205.10 (-5.16)	-162.746 (-4.00)	5.969 (209.49)	6.006 (204.76)
R-Square	.64	.66	.67	.69
SSE	166681786	156396051	85.5	81.2

* multiplied by LOTSZ.

TABLE III.3

Preliminary Regression Results for the
Linear and Semilog Functional Forms
With LOTSZ Interaction Terms
(t-ratios in parenthesis)

Variable	Linear		Semilog	
	Specification 1	Specification 2	Specification 1	Specification 2
AREA	.529 (22.02)	.500 (21.8)	.00037 (22.05)	.00035 (22.14)
BATH	66.39 (3.18)	94.98 (4.75)	.047 (3.21)	.069 (5.06)
AGE	1.718 (3.20)	.304 (.58)	.000095 (.25)	-.0012 (-3.03)
LOTSZ	8.10 (.12)	-35.23 (-.24)	-.0271 (-.56)	-.10 (-1.0)
FIRE	37.073 (2.23)	29.65 (1.88)	.077 (6.63)	.069 (6.43)
AIR	-6.40 (.31)	-16.12 (-.82)	-.010 (-.72)	-.018 (-1.36)
POOL	30.985 (1.32)	67.77 (3.03)	.012 (.75)	.043 (2.81)
VIEW	409.97 (10.60)	269.99 (6.62)	.203 (7.52)	.122 (4.38)
WHTCT	4.823 (5.46)	5.12 (5.01)	.006 (9.87)	.0071 (10.14)
WRKCT	-13.07 (-2.74)	4.32 (.77)	-.007 (-2.19)	.0014 (.36)
BEACH	16.88 (5.90)	4.76 (1.01)	.012 (6.21)	-.0043 (-1.34)
CBD	5.842 (1.99)	-7.67 (2.15)	.0024 (1.18)	-.009 (-3.52)
VIS	77.48 (6.90)	29.00 (1.41)	.060 (7.65)	.010 (.69)
TSP	-14.478 (-8.34)	-18.87 (9.28)	-.011 (9.17)	-.013 (-9.22)
CRME		-3738.24 (-3.5)		-4.05 (-5.56)
D1	-	-542.11 (-6.5)	-	-.49 (-8.6)
D2	-	-592.33 (-.6)	-	-.763 (-1.2)
03	-	-320.57 (-2.25)	-	-.004 (-.04)
VIEW-LOTSZ	1.70 (.25)	16.06 (1.77)	.0025 (.52)	.0007 (.11)

TABLE III.3
(Continued)

Preliminary Regression Results for the
Linear and Semilog Functional Forms
With LOTSZ Interaction Terms
(t-ratios in parenthesis)

Variable	Linear		Semilog	
	Specification 1	Specification 2	Specification 1	Specification 2
WHTCT-LOTSZ	-.163 (-.76)	.414 (1.55)	-.00005 (-.34)	.00008 (.45)
WRKCT-LOTSZ	3.484 (2.94)	-1.472 (-.78)	.0021 (2.53)	.0004 (.32)
BEACH-LOTSZ	-2.08 (-3.39)	-4.90 (-3.04)	-.0015 (-3.60)	-.0014 (-1.05)
CBD-LOTSZ	.225 (.20)	1.262 (.86)	.00065 (.84)	.001 (.97)
VIS-LOTSZ	8.73 (2.04)	-3.954 (-.48)	.0058 (1.93)	.0008 (.14)
TSP-LOTSZ	-1.14 (-2.34)	1.774 (2.50)	-.0005 (-1.52)	.001 (2.08)
CRME-LOTSZ		-20.27 (-2.19)		-.004 (-.62)
D1-LOTSZ	-	62.22 (1.82)	-	.069 (2.97)
D2-LOTSZ	-	57.85 (.6)		.039 (.59)
D3-LOTSZ	-	94.188 (2.98)	-	.010 (.46)
INTERCEPT	17.20 (.08)	1143.37 (3.26)	5.901 (41.42)	6.964 (29.06)
R-SQUARE	.71	.74	.74	.78
SSE	135459584	120281477	66.249	56.02

I

TABLE III.4

Effect of Lotsize Treatment on the
Property Value Effects of VIS and TSP Focus
Variables (t-values in parenthesis)

	Linear 1			Linear 2			Semilog 1			Semilog 2		
VIS	87.67 (14.14)		77.48 (6.90)	15.36 (1.52)		29.00 (1.41)	.066 (15.35)		.060 (7.65)	.0043 (.63)		.010 (.69)
VIS X LOTSIZE		.044 (13.30)	8.73 (2.04)		.0097 (.61)	-3.954 (.48)		.000033 (13.95)	.0058 (1.93)		.00002 (3.69)	.0008 (.14)
TSP	-16.42 (15.29)		-14.48 (8.34)	-14.99 (13.26)		-18.87 (9.28)	-.012 (15.63)		-.011 (9.17)	-.010 (13.51)		-.013 (9.22)
TSP X LOTSIZE		-.0045 (10.19)	-1.14 (2.34)		-.0047 (6.97)	1.774 (2.50)		-.0000027 (8.31)	-.0005 (1.52)		-.0000033 (6.85)	.001 (2.08)

Lotsize Mean = 1978.22

Price Mean = 1033.14

TABLE III.5

Partial Effects of VIS and TSP on
Property Values
(at means of Lotsize and Price)

	Linear 1			Linear 2			Semilog 1			Semilog 2		
	Alone	Interacted	Both	Alone	Interacted	Both	Alone	Interacted	Both	Alone	Interacted	Both
VIS	\$87.67	\$87.04	\$17,347	\$15.36	\$19.19	\$-7,793	\$68.19	\$67.44	\$11,854	\$4.44	\$40.88	\$1,645
TSP	-16.42	-8.90	-2,270	-14.99	- 9.30	+3,491	-12.40	-5.52	-1,033	-10.33	-6.74	2,030

IV. Measurement Error

A. Introduction

The issue of measurement error has a long history in econometric theory, although techniques attempting to correct for measurement error have rarely found their way into applied work in economics. This fact is not so much due to naivety on the part of researchers as it is to the difficulties in treating the problem. We do not offer a solution to the problem, but instead in the spirit of this report try to determine its importance.

We employ a variation of the methodology suggested by Klepper and Learner (1984) and first used in the hedonic literature by Atkinson and Crocker (1984). The goal of the analysis is to try to identify how important measurement error in any variable is with respect to the estimation of the coefficients of interest, with the idea being that better measurements would be necessary if the variability in parameter estimates was too great.

B. Bounding the parameters estimates

1. Theory

The Klepper-Leamer approach extends an idea of Frisch (1934) that bounds on ML parameter estimates can be obtained by "reverse" regressions. That is, the regression equation is solved for each variable potentially measured with error and the least squares fit is obtained. Then the coefficients of interest are solved out, and the minimum and maximum values over the separate regressions form the bound. When there are several variables (Frisch considered the simple regression case) the ML estimate is to be found in the "core" of the separate estimates only if the

separate estimates of the parameters lie in the same orthant. Otherwise the ML estimates could be any numbers, depending upon the correlation between the measured and the true variable. An example illustrates.

2. An example

Suppose we have a regression model with two exogenous variables. Suppressing the constant and the disturbance:

$$Y = B_1X_1 + B_2X_2$$

Solving in turn for X_1 and X_2 gives

$$x_1 = \frac{1}{\beta_1} y - \frac{\beta_2}{\beta_1} x_2$$

$$x_2 = \frac{1}{\beta_2} y - \frac{\beta_1}{\beta_2} x_1$$

Thus three regressions can be run. In the first regression we obtain estimates of B_1 , and multiply the negative of this number times the coefficient of X_2 to compute our estimate of B_2 . The technique follows analogously in the third regression. The results produce a table such as the following (numerical entries are hypothetical):

Independent Variables	Dependent Variable	Y	X_1	X_2
	X_1	-2.1	-1.7	-3.7
	X_2	4.3	2.0	1.1

Estimates Coefficients, Original and Reverse Regressions

In this example, all three parameter estimate vectors lie in the same orthant (namely, the second), so that the ML estimator also lies in the second quadrant, and is bounded as indicated in Figure IV.1.

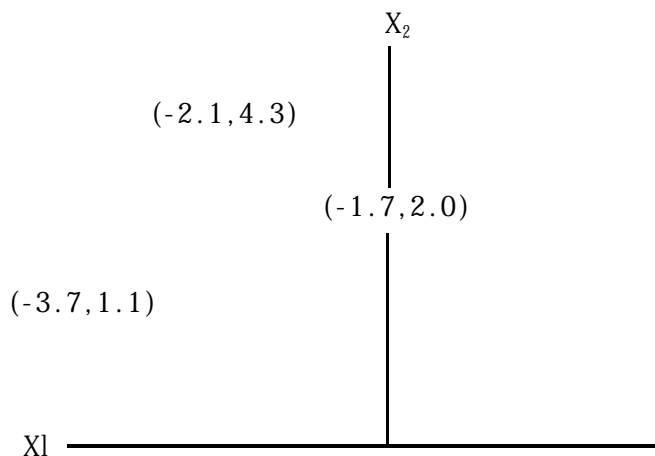


Figure IV.1

3. The methodology applied to our data

If it is suspected that measurement error may be present in any of the variables in a hedonic price study, the first step then is to run the reverse regressions and examine the signs of the coefficients (for the signs determine the quadrant).

As can be seen in Table IV. 1, the results from the reverse regressions with the Los Angeles data set indicate that the coefficients for the focus variables VIS and TSP are unbounded in both specifications. The variables assumed to be measured with error (WHTCT, WRKCT, BEACH, LOTSZ, VIS, TSP, and CRME) were chosen primarily for illustration. However, with the exception of LOTSZ, they represent the neighborhood, location, and environmental variables. These variables are likely measured with error when compared to the site specific variables. In fact, the neighborhood quality variables can be considered proxy measures and are, therefore, measured with error by definition. BEACH is distance to the nearest-beach in miles, but this attribute may be more accurately measured by time to nearest beach. LOTSZ is average lot size based on community housing

density. Clearly, there is measurement error here. For communities with parks, golf courses, etc., the LOTSZ variable increases, even though the actual lot size may remain unchanged. Similar types of arguments can be formulated for CRME. VIS and TPS are the environmental quality proxy variables. One purpose here is to determine if more accurate measurement of these variables would significantly improve the estimates of the hedonic prices for environmental quality.

In Table IV.1, LOTSZ in specification 1 is the only variable that does not change sign. Thus contrary to the example presented in Figure IV.1 no bound can be placed on our parameter estimates.

TABLE IV.1

Signs of the Coefficients on the Variables Possibly
Measured with Error by Alternative Dependent
Variables -- Reverse Regression Results

(Specification 1)

Dependent Variable	P	WHTCT	WRKCT	BEACH	LOTSZ	VIS	TSP
WHTCT	+	+	-	+	+	-	+
WRKCT	-	+	-	-	-	-	+
BEACH	+	+	+	+	+	+	+
LOTSZ	+	+	+	+	+	+	+
VIS	+	-	+	+	+	+	+
TSP	-	-	+	-	-	-	-

(Specification 2)

Dependent Variable	P	WHTCT	WRKCT	BEACH	LOTSZ	VIS	TSP	CRIME
WHTCT	+	+	+	+			+	
WRKCT	+	+	+				+	
BEACH			+			+	+	
LOTSZ	+			+	+	+		+
VIS	+				+	+	+	+
TSP				+	+			+
CRIME		+	+				+	

Therefore it appears that, for our data, the possibility of measurement error leaves open even the question of the direction of the effect of our variables.

C. Minimum correlations

1. Theory

By adding more structure to the model we may hope to answer the following question: what is the minimum correlation between actual and true variables necessary before at least the signs of the coefficients are unambiguous? Assume the relationship between the measured and the true variables is given by

$$X_{ij} = X_{ij}^* + v_{ij}$$

where X_{ij} is the measured value of the j -th observation on the i -th explanatory variable, X_{ij}^* is its true counterpart, and v_{ij} is white noise.

Specifying the normal errors in variables model we have

$$\begin{pmatrix} X_{ij}^* \\ X_{ij} \\ V_{ij} \end{pmatrix} \sim N_3 \begin{pmatrix} \mu_1 & \delta_1^2 + \delta_{v_1}^2 & \delta_1^2 \\ \mu_1 & \delta_1^2 & 0 \\ 0 & \delta_{v_1}^2 & \delta_v^2 \end{pmatrix}$$

Thus X_{ij}^* is drawn from a normal distribution with mean μ_1 , V_{ij} is added to it to form X_{ij} . The V_{ij} are uncorrelated with the X_{ij}^* . This

means the squared correlation coefficient between X_{ij}^* and X_{ij} is

$$P_i^2 = \frac{\text{COV}(x_{ij}, x_{ij}^*)}{V(x_{ij}) V(x_{ij}^*)} = \frac{(s_i^2)^2}{(s_i^2 + s_{v_i}^2) s_i^2} = \frac{s_i^2}{s_i^2 + s_{v_i}^2}$$

If this correlation is known, the ML estimator is

$\tilde{B} = (X'X - E)^{-1}X'y$, where E is a diagonal matrix whose i -th diagonal element is $1 - p_i^2$. Note that as the p_i^2 approaches one E approaches the zero matrix and \tilde{B} becomes the least squares estimator $(X'X)^{-1}X'y$. This is as it should be, since $p_i^2 = 1$ implies that the observed variable x_{ij} measures the true variable x_{ij}^* without error and, under normality, ML and least squares coincide.

It is of course impossible to know (or even estimate) p_i^2 , since x_{ij}^* is not observed. The usual approach to estimation is to try to find an instrumental variable that is highly correlated with x_{ij}^* but uncorrelated with v_{ij} . This is a difficult task, as with any instrumental variable estimation. Since we are examining parameter variability here, we have the luxury of specifying potential values for the p_i^2 and examining the parameter-estimates of interest.

2. The methodology applied to our data.

Focusing on WHTCR, WRKCT, BEACH, LOTSZ, VISZ, TSP, and CRME, estimates for 1 are presented in Table IV.2, assuming $p_i^2 = <1$ for all x_i that are assumed to be measured with error and with $p_i^2 = 1$ for all i and x_i that are assumed to be measured accurately; AREA, BATH, AGE, FIRE, AIR, POOL, VIEW, CBD, D1, D2, AND D3. The estimates vary substantially, even when p_i^2 is 95. This indicates that any inferences drawn from the hedonic estimates should be qualified considerably.

Our prior expectations for VIS and TSP are a positive influence for VIS and negative influence for TSP. This prior is violated in both specifications. However, it is difficult to analyze the importance of measurement error to the estimates on VIS and TSP in Table IV.2, since all the p_i^2 's change together. Therefore, estimates for the coefficients on VIS and TSP were obtained by changing only one p_i^2 at a time and assuming the other variables are measured accurately. These are presented in Table IV.3 (specification 1) and Table IV.4 (specification 2).

In specification 1, the estimates for the coefficient on VIS and TSP are quite stable when WHTCT, WRKCT, and LOTSZ are measured with error. They seem more sensitive to measurement error in BEACH, VIS, and TSP. Although TSP remains quite constant for different p_i^2 's on VIS (Fifth row of Table IV.3), its value is less than half the value obtained with no measurement error. Similarly, VIS is quite stable as TSP's correlation changes, but is much smaller than the .066 measure reported in Section 2. The conclusions are basically the same for specification 2. However, in specification 2 WRKCT seems to cause VIS to fall quite substantially.

D. Conclusions

An important observation from Tables IV.3 and IV.4 is that our priors (in terms of expected signs on estimated coefficients) are realized when WHTCHT, WRKCT, BEACH, LOTSZ, and CRME are presumed to be measured with error. They are not when either VIS or TSP are modeled as measured with error. This may indicate that the emphasis for future environmental economics research should concentrate primarily on measuring the environmental quality variables.

TABLE IV.2

Estimated Coefficients for the Variables
Possibly Measured with Error by
Various Values for RHO. Semilog Form

(Specification 1)

Variable	rho = .5	rho = .6	rho = .7	rho = .8	rho = .9	rho = .95	rho = .97	rho = .99
WHTCT	-.00034	-.00046	-.00075	-.00012	-.0015	-.004	-.0073	-.0185
WRKCT	-.0001	.00018	.00048	-.0013	-.00073	-.0035	.0474	-.0233
BEACH	.0089	.0142	.0354	.0797	-.020	-.018	-.013	-.0211
LOTSZ	-.0000006	-.00000087	-.0000018	-.000003	.0000006	.00000016	-.0000002	.000000
VISZ	-.003	-.0045	-.009	-.010	-.006	-.028	.128	-.036
TSP	.0003	.0004	.0009	-.0013	.00011	.00012	.013	.0154

(Specification 2)

Variable	rho = .5	rho = .6	rho = .7	rho = .8	rho = .9	rho = .95	rho = .97	rho = .99
WHTCT	-.00035	-.00045	-.00064	.0012	-.0013	-.0041	-.0116	.0046
WRKCT	-.000016	-.00005	-.00016	-.0010	.0026	.0045	.0121	-.0403
BEACH	.006	.0085	.0143	.045	-.039	-.020	-.0171	-.0619
LOTSZ	-.0000004	.00000027	.0000003	.000005	-.000014	-.000021	.006	.00037
VISZ	-.0017	-.0024	-.0042	-.014	.014	.0102	.011	.051
TSP	.0003	.0004	.00059	.0013	.0003	.002	.0048	.0022
CRME	-.032	-.05	-.118	-.667	1.526	2.164	4.573	-39.147

TABLE IV.3

Estimated Coefficients for VIS and TSP when
WHTCT, WRKCT, BEACH, LOTSZ, VIS, and TSP are Assumed to
be Measured with Error by Various Values of Rho.

(Specification 1)

Variables Measured with Error	rho				
	.5	.6	.7	.8	.9
WHTCT	.075 -.013	.075 -.013	.075 -.013	.075 -.013	.076 -.013
WRKCT	.066 -.012	.066 -.012	.066 -.012	.066 -.012	.066 -.012
BEACH	.050 -.009	.048 -.008	.046 -.008	.038 -.006	.196 -.036
LOTSZ	.067 -.012	.067 -.012	.066 -.012	.066 -.012	.066 -.012
VIS	-.002 -.005	-.002 -.005	-.0034 -.005	-.005 -.005	-.011 -.005
TSP	.030 .0002	.030 .0002	.030 .0003	.029 .0004	.027 .0009

NOTE: In each row of the Table, the other variables are considered to be measured accurately.

TABLE IV.4

Estimated Coefficients for VIS and TSP
When WHTCT, WRKCT, BEACH, LOTSZ, VIS, TSP
and CRME, are assumed to be Measured with Error by
Various Values of Rho

(Specification 2)

Variables Measured with Error	rho				
	.5	.6	.7	.8	.9
WHTCT	.042 -.012	.042 -.012	.043 -.012	.044 -.012	.048 -.012
WRKCT	.0067 -.010	.0067 -.010	.0067 -.010	.0068 -.010	.070 -.010
BEACH	.028 -.013	.028 -.013	.029 -.013	.031 -.013	.038 -.014
LOTSZ	.0034 -.011	.0034 -.011	.0034 -.011	.0034 -.011	.0032 -.011
VIS	-.000043 -.010	-.00005 -.010	-.00007 -.010	-.0001 -.010	-.0002 -.010
TSP	-.043 .00012	.043 .00015	.043 .0002	.044 .0003	.045 .0006
CRME	.004 -.011	.004 -.011	.004 -.011	.004 -.011	.004 -.011

NOTE: In each row of the table, the other variables are, considered to be measured accurately.

V. Effect of Functional Form

A. Introduction

Economic theory usually has little to say about correct functional form. Often an appeal is made to the Taylor series expansion to justify linearity, but of course a first-order Taylor series expansion is useful only near the point of approximation. In some cases (perhaps, for example, demand theory) a long history of proportional response implies a constant-elasticity specification, but this is still an empirical observation. For the situation of hedonic gradients, which are by theory already in reduced form and hence are solutions of several equations, even less can be presumed. It would seem then unreasonable to impose a priori a structure on the data. In nearly every study to date this is in fact what is done, or a non-systematic search over a few functional forms is made and only the result conforming most closely to the priors of the investigator is reported.

This is surprising considering the energy economists have put into the development of alternative functional forms for production and cost, as well as the extensive work in the statistics literature on variable transformations. Halvorsen and Pollakowski (1981) nicely combine these ideas to suggest a model sufficiently general to include many of the most popular specifications (including linear, log-linear, log-log, quadratic and translog). Since all of the functional forms are included in the most general functional form, called the quadratic BOX-Cox, conventional tests of hypotheses are available (e.g., likelihood ratio tests). Although estimation is by maximum likelihood, the likelihood function condenses considerably so that the computational burden is not onerous.

B. The quadratic Box-Cox model

1. Specification

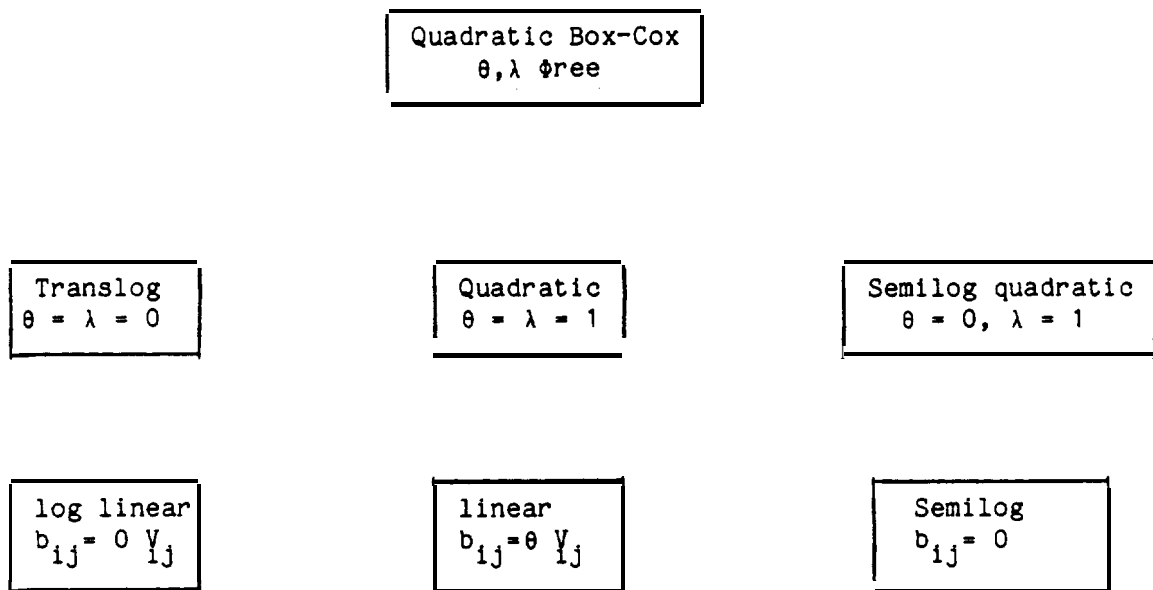
The quadratic Box-Cox model is

$$P(0) = a_0 + \sum_{i=1}^m a_i X_i(\wedge) + 1/2 \sum_{i=1}^m \sum_{j=1}^m b_{ij} x_i(\wedge) x_j(\wedge)$$

$$p(0) = \begin{cases} (P_{\text{Inp}}^0 - 1)/0 & 0 \neq 0 \\ 0 & 0 = 0 \end{cases}$$

$$X_i(\wedge) = \begin{cases} (x_i^\wedge - 1)/\wedge & \wedge \neq 0 \\ \ln X_i & \wedge = 0 \end{cases}$$

We consider three pairs of nested specifications, shown by the “tree” diagram, with the necessary parameter restrictions:



2. Estimation

A two-step technique is implemented for maximizing the likelihood function. First, values for θ and λ are chosen and the data are transformed. Then the concentrated log-likelihood is evaluated:

$$L(\theta, \lambda) = -\frac{1}{2} n \log \sigma^2(\theta, \lambda) + (0-1) \sum_{i=1}^n \log P_i$$

where n is the sample size, $\sigma^2(\theta, \lambda)$ is the ordinary least square estimator of the variance of the transformed data, and P_i is the i -th observation on price. A search over a two-dimensional grid for the largest value of $L(\theta, \lambda)$ produces the maximum likelihood estimates of θ, λ , and the α 's and b_{ij} 's (An even more general model is possible if it is specified that there is a different transformation for different exogeneous variables, although this would greatly increase computer costs as the search would be conducted over an $(M+1)$ -dimensional grid).

3. Hypothesis testing

Having obtained the maximized value of the log-likelihood function and resulting parameter estimates, it is a straightforward exercise to test the hypothesis concerning more restrictive functional forms. To do so, θ and λ are set to their respective values under the null hypothesis, α_i 's and b_{ij} 's are again found, and the new value of the likelihood is computed. Then minus twice the difference between the constrained log-likelihood and the maximum value of the log-likelihood has, asymptotically, the (chi-squared) distribution with two degrees of freedom if the null hypothesis is true. The most restrictive specifications (log-linear, linear and semi-log) may be tested either unconditionally or against their parent members (translog, quadratic and semilog quadratic, respectively).

C. Empirical results

1. Specification 1

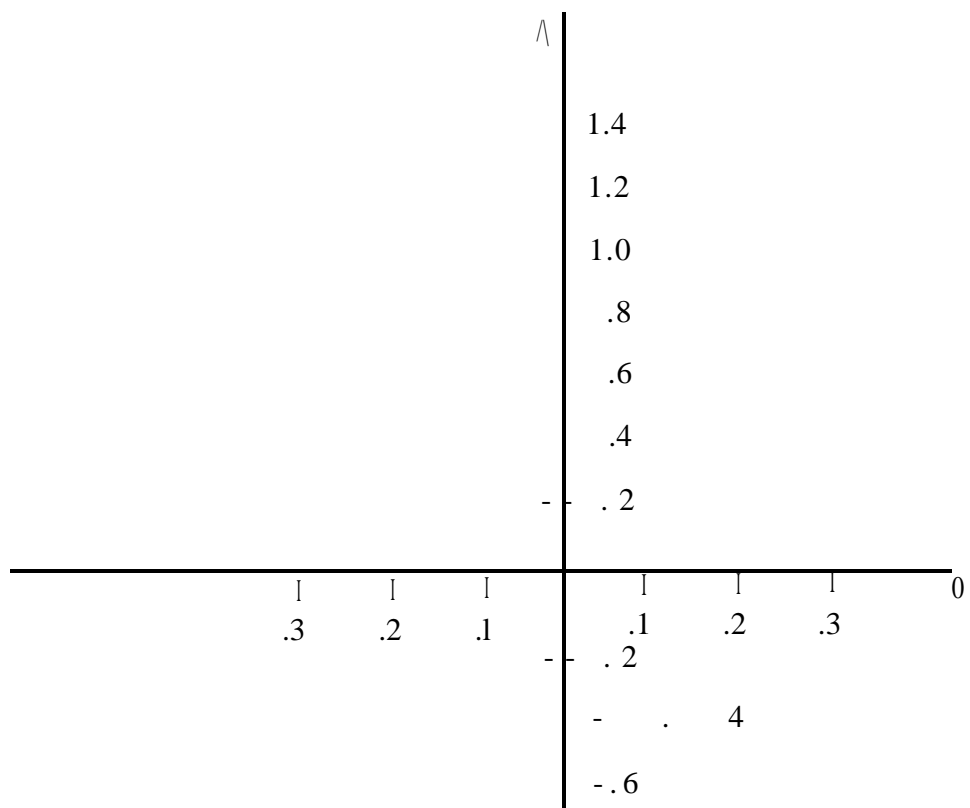
As indicated above the initial step in the empirical analysis is to estimate the unknown parameters for the quadratic Box-Cox form using the maximum likelihood procedure. In the first specification the optimum optimum is found where $\theta = - .10$ and $\lambda = 1.10$. This yields a value for the likelihood function $L(\theta, \lambda)$ of -7328.14 . This result can be utilized to determine a confidence region around the estimates θ, λ . The $100 (1-\alpha\%)$ confidence region consists of all points (θ^*, λ^*) which satisfy the inequality:

$$L_{\max}(\theta, \lambda) - L_{\max}(\theta^*, \lambda^*) < 1/2 \chi_{\kappa}^2(\alpha)$$

where κ is the number of restrictions and α is the significance level. In the case where both θ and λ are restricted then $\kappa = 2$ and the χ^2 value for the 99 percent confidence region is 4.605. The confidence region pertaining to $\alpha = .01$ for the initial specification is shown in Figure V.1. As is evident, the value of the log-likelihood is slightly more sensitive to the transformation parameter for the dependent variable (θ) than to the transformation parameter of the independent variable (λ).

Given the estimated parameters for the quadratic Box-Cox a variety of other functional forms can be estimated by restricting the choice over θ and λ . The values of (θ, λ) for each form are: translog (0,0) quadratic (1,1) and semi-log quadratic (0,1). The values for the log-likelihood, excluding a constant, the the translog, quadratic and semi-log quadratic are -7341.36 , -7859.77 and -7334.02 , respectively. Calculation of the relevant χ^2 statistic, or examination of Figure V.1, in which the values of θ and λ corresponding to these forms are found to lie outside the 99 percent confidence interval, implies that all of these forms can be

FIGURE V.1
99% Confidence Region



rejected at the .01 level. In addition, rejection of their parent members suggests that the lower order forms (log-linear, linear and semi-log) can also be rejected at the .01 level .*

2. Specification 2

The results of the functional form analysis for our second specification are quite similar. The optimum optimum for the quadratic **Box-Cox occurs at $\theta = -.13$ and $\lambda = -.35$** . The value of the log-likelihood function is -7243.26. The 99 percent confidence region for specification two is presented in Figure V.2. As in the previous case, the values of θ and λ corresponding to the other estimated forms lie outside the confidence region suggesting rejection of the forms at the .01 level.

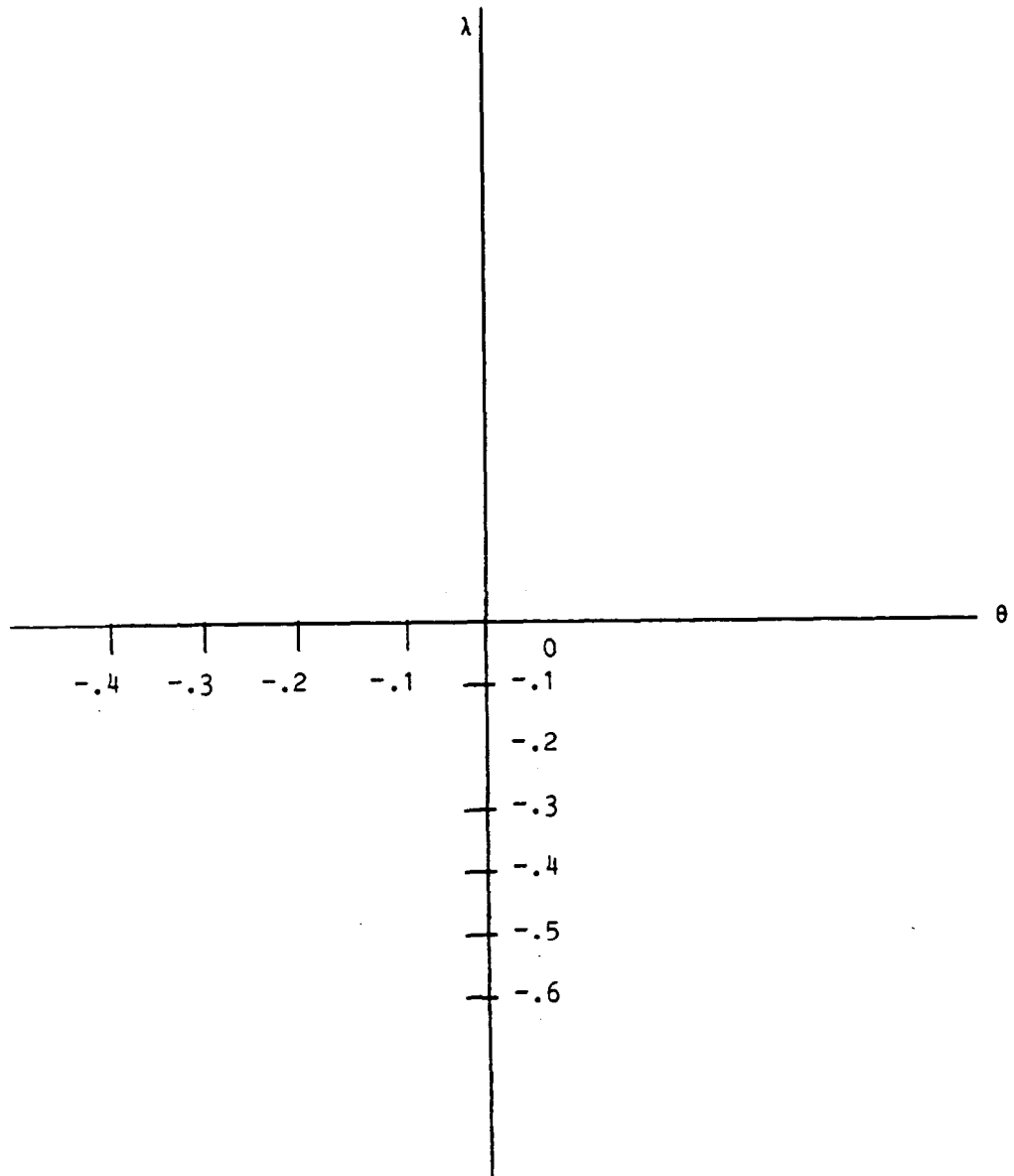
The empirical analysis to this point implies that most commonly used functional forms can be rejected on a statistical basis. However, the relative impact of functional form variation on benefit estimation has yet to be investigated. Thus, if the estimated hedonic prices are relatively insensitive to functional form then commonly used forms may provide relatively precise benefit estimates. In this case, the more complex forms may be unnecessary.

The relative sensitivity of the estimated hedonic prices for visibility and total suspended particulates (TSP) is illustrated in Tables V.3 and V.4 for specifications 1 and 2, respectively. The hedonic prices for visibility and TPS were calculated assuming that all characteristics were assigned their mean values since the price of any characteristic depends upon the levels of all other characteristics.

*The values, excluding a constant, of the log-likelihood for the log-linear, linear and semi-log are -7543.71, -8142.78, and -7528.71, respectively. Either an unconditional test or a test against their parent members suggests rejection.

FIGURE V.2

99 Percent Confidence Region



In Table V.3 the hedonic prices for visibility and TSP are relatively stable. Visibility ranges from approximately 60 to 10 whereas TSP ranges from approximately 7 to 16. The ratios of highest to lowest mean values are then 1.67 and 2.33 for the two amenity variables, respectively. In each case, the quadratic Box-Cox yields the lowest mean price indicating that previous analyses provided overestimates of benefits of environmental improvements. *

However, the results in Table V.4 completely contradict those in Table V.3; that is, for the second specification the mean hedonic prices are much more sensitive. The visibility and TSP mean prices vary by ratios of 177/1 and 659.7/1, respectively. Even if the extremely volatile quadratic Box-Cox functional form is excluded the mean visibility hedonic price varies by a ratio of 9.65/1. In addition, the results for specification 2 indicate that the quadratic Box-Cox produces the largest benefit estimates. Thus, analyses that utilize restricted functional forms would underestimate benefits of environmental improvements.

D. Conclusions

The following conclusions concerning the importance of functional form are suggested by the preceding empirical analysis. First, the more general functional forms significantly outperform the restrictive forms. Second, the effect of functional form on the mean hedonic prices depends primarily upon the model specification. Thus, the results are inconsistent, emphasizing again the importance of variable selection in specification of the hedonic price equation. It should also be noted that

* Although these hedonic prices are not strictly interpretable as benefits, it is usually the case that high (hedonic) prices produce higher benefit estimates. See Murdoch and Thayer (1984).

the hedonic prices may be even more sensitive than portrayed in this analysis. That is Murdoch and Thayer (1983), have with a similar data set, found that homes with characteristics significantly different from the mean home are much more sensitive to functional form variation.

TABLE V.3

Summary Statistics for the Predicted
Hedonic Price of Visibility and TSP
by Functional Form of the Hedonic
Price Equation (Specification 1)

Functional Form	Variable	Mean	Standard Deviation	Minimum	Maximum
Quadratic Box-Cox	VIS	60.08	87.80	-202.7	711.8
	TSP	7.04	28.04	-14.57	681.2
Translog	VIS	97.59	85.61	-48.99	789.17
	TSP	9.87	12.11	-21.13	102.98
Quadratic	VIS	100.52	116.68	145.68	383.85
	TSP	11.57	44.53	-24.19	975.46
Semilog	VIS	78.94	107.55	-220.3	925.7
Quadratic	TSP	10.48	31.28	-19.52	726.6
Log-linear	VIS	70.61	27.74	18.35	302.0
	TSP	11.58	6.10	2.39	57.85
Linear	VIS	87.66	0	87.66	87.66
	TSP	16.42	0	16.42	16.42
Semi log	VIS	66.76	35.22	27.59	592.79
	TSP	11.78	6.21	4.87	104.55

NOTE: prices stated in \$100's. TSP is for a reduction of TSP.

TABLE V.4

Summary Statistics for the Predicted Hedonic
Prices of Visibility and TSP by Functional
Form of the Hedonic Price Equation
(Specification 2)

Functional Form	Variable	Mean	Standard Deviation	Minimum	Maximum
Quadratic Box-Cox	VIS	781.29	1824.5	-15143.4	14606
	TSP	5370	5045	-7512	35769
Translog	VIS	31.86	65.62	-433.31	581.30
	TSP	8.14	11.28	- 31.65	76.8
Quadratic	VIS	42.60	173.9	-151.7	3496.0
	TSP	11.93	30.10	-25.98	611.5
Semilog Quadratic	VIS	30.12	242.94	-124.66	5901.56
	TSP	17.43	52.81	- 13.3	1284.5
Log-linear	VIS	10.36	4.24	2.78	41.69
	TSP	10.58	5.78	2.25	58.26
Linear	VIS	15.36	0	15.36	15.36
	TSP	15.00	0	15.00	15.36
Semilog	VIS	4.41	2.33	1.49	35.41
	TSP	10.52	5.55	3.56	84.45

NOTE: prices stated in \$100's. TSP is for a reduction of TSP.

VI. Robust estimation

A. Introduction

There is a growing awareness in the applied economic literature that the heavy reliance on the assumption of normality may seriously bias parameter estimates if in fact the model is misspecified. The general feeling seems to be that the normal distribution has too little weight in its tails, so that great weight is placed on outlying observations. Recent work in the area include attempts at the detection of influenct observations (Belsley, Kuh, and Welsch, 1980) nonparametric maximum likelihood estimation (Manake, 1975; Cosslett, 1983), and alternative fitting criteria that minimize the effect of outliers (Koenker and Basset, 1982; Guilkey and Waldman, 1985. It is this last approach that we follow here.

The assumption of normally distributed regression disturbances may be loosely justified by appealing to a variant of the central limit theorem. This appeal requires as preconditions a correctly specified model, with no omitted variables, and as components of the disturbance many small, independently distributed random variables uncorrelated with the explanatory variables in the model. If these preconditions are not met, as is likely, disturbances will not be normally distributed and consequently least square estimation will no longer be optimal.

The case is often made that least squares is robust to misspecification. This is true to an extent, but there is ample evidence (see Koenker, 1982,) that serious biases can also result.

B. The minimum absolute deviation estimator

In an attempt to assess the importance of the usual least squares estimation methodology applied to most studies of hedonic markets, we reestimated our basic specifications employing the fitting criterion of "minimum absolute deviations" (MAD). Algebraically, for the model $Y =$

$\beta'x_i + \epsilon_i$, $i = 1, \dots, n$, the criterion is to choose β such that

$\sum_{i=1}^n |y_i - \beta'x_i|$ is minimized. The idea is that outlying observations are given more weight, as they should be, but only in proportion to their distance from the center rather than the square of that distance. The results are presented in Table ____.

VII. Concluding remarks

In this section we summarize our findings. After introducing the focus of our study and reviewing past work, we detailed the characteristics and discussed the applicability of our chosen data set (sections I and II, respectively).

With respect to variable selections, we found a curious dual result: the coefficients of one measure of air quality, an index of visibility (VIS) were quite (can't read_) with respect to which subset of doubtful variables were included in the analysis, while the coefficients of another measure total suspended particulates (TSP), were remarkably stable. The coefficients of VIS were variously negative and significant, negative and

insignificant, and positive, both insignificant and significant. The coefficient estimates for TSP were contained in a narrow interval, especially when VIS was also included in the equation. We may speculate here that although homeowners obviously prefer better visibility, they are more strongly and systematically worried about the long run health and soiling implications of pollution. [Note to co-authors: need correlation matrix results to determine econometrics issue of collinearity and how it affects these estimates].

Unfortunately the positive results on the impact of TSP on property values are cast into doubt if the issue of the state of equilibrium of the land market is considered. This was demonstrated by recalculating results from regressions incorporating interactions of lotsize and air quality. Then, even within the same specification (functional form and choice included variables) the environmental evaluations were variously positive and negative depending upon the treatment of lotsize.

In section IV we examined the issue of measurement error. Here we found that potential measurement error in the control variables did not affect the overall results as seriously as potential measurement error in the focus variables (TSP and VIS).

In Section V we examined the effect of functional form. We found that the more general functional forms significantly outperform the more restrictive functional forms. However, even within the most general functional form (the quadratic Box-Cox), the choice of included variables greatly affected the results.

Finally, we employed a more robust estimator (the minimum absolute deviation estimator) in an attempt to find an alternative to least squares.

BIBLIOGRAPHY

- Anderson, J.E. "Ridge Estimation of House Value Determinants," J. of Urban Economics, Vol. 9, No. 3, May 1981, pp. 286-297.
- Anderson, J.E., and T.D. Crocker. "Air Pollution and Residential Property Values," Urban Studies Vol. 8, No. 3, October 1971, pp. 171-180.
- Atkinson, Scott E. and Thomas D. Crocker. "Specification Uncertainty and Measurement Error Intolerance in Hedonic Property Value Studies," delivered at the NBER-NSF Seminar on Bayesian Inferences in Econometrics, May 4-5, 1984.
- Bartik, J.J. and V.K. Smith. "Urban Amenities and Public Policy," Vanderbilt University, 1984.
- Belsley, D., E. Kuh, and R. Welsch. "Regression Diagnostics, Identifying Influential Data and Sources of Collinearity," Wiley, New York, 1980.
- Berger, et al. (EPA Report).
- Blomquist, G. and L. Worley. "Hedonic Prices, Demands for Urban Housing Amenities, and Benefit Estimates," J. of Urban Economics, Vol. 9, No. 2, March 1981, pp. 212-221.
- Brookshire, D., M. Thayer, W. Schulze, and R. d'Arge. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," American Economic Review, March 1982, 72.
- Brown, G.M., Jr. and H.O. Pollakowski. "Economic Valuation of Shoreline," The Review of Economics and Statistics, Vol. LIX, No. 3, August 1977, (North-Holland Publishing Co.), pp. 272-278.
- Cosslett, S.R. "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51, 1983, pp. 765-782.

- Frisch, R. "Statistical Confluence Analysis by Means of Complete Regression Systems," University Institute of Economics, Oslo, Norway, 1934.
- Goodman, A.C. "Hedonic Prices, Price Indices and Housing Markets," Journal of Urban Economics, Vol. 5, 1978 pp. 471-84.
- Craves, P.E. and T.A. Knapp. "Hedonic Analysis in a Spatial Context," The Economic Record, 1985, forthcoming December.
- Guilkey, D. and D. Waldman. "A Trimmed Estimator for the Probit Model," 1985, Manuscript.
- Halvorsen, R. and H.O. Pollakowski. "Choice of Functional Form for Hedonic Price Equations," J. of Urban Economics, Vol. 10, No. 1, July 1981, pp. 37-49.
- Horowitz, J.L. "Bidding Models of Housing Markets," Unpublished paper. U. of Iowa, March 1984.
- Haurin, D.A. "Urban Structure, Wage Rates, and Regional Amenities," in D. Diamond and G. Tolley (eds) The Economics of Urban Amenities New York: Academic Press, 1982.
- Kadiyala, K.R. Regression with Non-Gaussian Stable Disturbances: Some Sampling Results," Econometrica, Vol. 40, No. 4, July 1972, pp. 719-722.
- Kish, L. and J.B. Lansing. "Response Errors in Estimating the Value of Homes," Journal of the American Statistical Association, Vol. 49, 1954, pp. 520-532.
- Koenker, R. "Robust Estimation in Econometrics," Econometric Reviews.
- Koenker, R. and C. Bassett. "Robust Tests For Heteroscedasticity Based on Regression Quantities," Econometrica 50 (no. 1), 1982, pp. 43-62.

- Klepper, S. and E.E. Learner. "Consistent Sets of Estimates for Regressions with Errors in All Variables," Econometrica, Vol. 52, No. 1, January 1984, pp. 163-183.
- Manski, C. "Maximum Score Estimation of the Stochastic Utility Model of Choice," Journal of Econometrics, 3, 1975, pp. 205-228.
- Murdock and Thayer, "Choice Among Distributions of Environmental Quality," Mimeo. Correspondence: J. Murdoch, Department of Economics and Finance, Northeast Louisiana University, Monroe, Louisiana, 71209, 1985.
- Nelson, J.P. "Residential Choice, Hedonic Prices, and the Demand for Urban Air Quality," J. of Urban Economics, Vol. 5, No. 3, July 1978, pp. 357-369.
- Ridker, R. and J.A. Henning. "The Determinants of Residential Property Values with Special Reference to Air Pollution," Review of Economics and Statistics, Vol. 49, No. 2, May 1967, pp. 246-257.
- Roback, J. "Wages, Rents, and the Quality of Life," Journal of Political Economy, Vol. 90, 1982, pp. 1257-78.
- Rosen, S. "Wage-Based Indexes of Urban Quality of Life," in P. Mieszkowski and M. Straszheim (eds) Current Issues in Urban Economics, Baltimore: Johns Hopkins Press, 1979.
- Rosen, S. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, Vol. 82, No. 1, January 1974, pp. 34-55.
- Trijonis, J., J. Murdoch, M. Thayer, and R. Hageman. "The Benefits of Visibility Improvements," report to the California Air Resources Board, 1984.

Wieand, Kenneth F. "Air Pollution and Property Values: A Study of the St. Louis Area." Journal of Regional Science, Vol 13, No. 1 (April 1973), pp. 91-95.

Zerbe, R.O., Jr. "The Economics of Air Pollution: A Cost-Benefit Approach." Toronto: Ontario Department of Public Health, (July) 1969.